A misspecification test for the higher order co-moments of the factor model

Lu, Wanbo; Yang, Dong; Boudt, Kris

published in
Statistics
2019

DOI (link to publisher)
10.1080/02331888.2019.1566907

document version
Publisher's PDF, also known as Version of record

Statistics
A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

A misspecification test for the higher order co-moments of the factor model

Wanbo Lu, Dong Yang & Kris Boudt

To cite this article: Wanbo Lu, Dong Yang & Kris Boudt (2019) A misspecification test for the higher order co-moments of the factor model, Statistics, 53:3, 471-488, DOI: 10.1080/02331888.2019.1566907

To link to this article: https://doi.org/10.1080/02331888.2019.1566907

STATISTICS
2019, VOL. 53, NO. 3, 471–488
https://doi.org/10.1080/02331888.2019.1566907

A misspecification test for the higher order co-moments of the factor model

Wanbo Lu[a], Dong Yang[a,b] and Kris Boudt[b,c]

[a]School of Statistics, Southwestern University of Finance and Economics, Chengdu, People's Republic of China; [b]Solvay Business School, Vrije Universiteit Brussel, Brussel, Belgium; [c]School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

ABSTRACT

The traditional estimation of higher order co-moments of non-normal random variables by the sample analog of the expectation faces a curse of dimensionality, as the number of parameters increases steeply when the dimension increases. Imposing a factor structure on the process solves this problem; however, it leads to the challenging task of selecting an appropriate factor model. This paper contributes by proposing a test that exploits the following feature: when the factor model is correctly specified, the higher order co-moments of the unexplained return variation are sparse. It recommends a general to specific approach for selecting the factor model by choosing the most parsimonious specification for which the sparsity assumption is satisfied. This approach uses a Wald or Gumbel test statistic for testing the joint statistical significance of the co-moments that are zero when the factor model is correctly specified. The asymptotic distribution of the test is derived. An extensive simulation study confirms the good finite sample properties of the approach. This paper illustrates the practical usefulness of factor selection on daily returns of random subsets of S&P 100 constituents.

ARTICLE HISTORY
Received 20 January 2018
Accepted 6 January 2019

KEYWORDS
Factor models; higher order co-moments; sparsity; curse of dimensionality; diagnostic test

AMS SUBJECT CLASSIFICATIONS
62F03; 91B25

1. Introduction

Accurate estimates of higher order co-moments are needed by decision makers with preferences about the skewness and kurtosis of the non-normally distributed payoff associated with their decision. This is particularly relevant in financial portfolio decisions where the stylized fact of non-normality of asset returns leads to the presence of skewness and kurtosis in portfolio returns. Given the multivariate nature of portfolio decisions, higher order co-moment estimates are an important input for the estimation of portfolio risk under the Cornish–Fisher expansion (see, e.g., [1,2]), the construction of mean-variance-skewness-kurtosis efficient portfolios (see, e.g., [3,4]), and the optimization of the second order Taylor expansion of the expected utility function (see, e.g., [5,6]).

CONTACT Wanbo Lu luwb@swufe.edu.cn School of Statistics, Southwestern University of Finance and Economics, Chengdu, People's Republic of China
Part of this work was done while the first author was visiting the Solvay Business School at Vrije Universiteit Brussel in the summer of 2017.

This paper aims to contribute to fill the gap between the theoretical superiority of higher order co-moment-based investment decisions and the practical challenge of estimating these higher order co-moments. For the latter, Martellini and Ziemann [6] and Boudt et al. [7] proposed estimators, which avoid the curse of dimensionality, by imposing a factor model structure on the return generating process. Among other things, this implies that the co-moments of the unexplained return variation are sparse. We intend to examine the validity of the proposed factor model by testing the sparsity of these co-moment matrices.

While the use of factor models is widespread in terms of modeling expected returns and estimating the covariance matrix, the estimation of the coskewness and cokurtosis matrix under the framework of a factor model is relatively recent (see, e.g., [6,7]). In previous literature, higher order co-moments are usually estimated by their sample analogues. Compared to the so-called 'restricted' higher order co-moments, which assume a specific return generating process such as a factor model, the unrestricted estimators have the advantage of being consistent, irrespective of the underlying return generating process. Their drawback is that there is a curse of dimensionality in terms of a steep increase in the number of parameters to estimate, when the dimension of the multivariate return series increases. This curse of dimensionality makes the unrestricted estimators of the first four (co)moments almost infeasible for moderately large dimensions (see, e.g., [8]).

The use of factor models for estimating the co-moments of stock returns thus solves the curse of dimensionality problem. However, it raises the problem of selecting an appropriate factor model. For analyzing stock returns, many candidate factor models exist. Martellini and Ziemann [6] used the single factor market model, while Boudt et al. [7] used statistical factors for modeling higher order co-moments. For modeling the mean and covariance matrix, the three-factor and five-factor models of Fama and French (see, e.g., [9,10]) are also popular.

Several approaches exist to select factors in a mean-variance setting. Cragg and Donald [11] and Bai and Ng [12] considered the use of information criteria, penalized regressions, and goodness of fit statistics. In the case of statistical factors, the scree plot is typically used to determine the number of principal components. Alternatively, tests involving the eigenvalues of the sample covariance matrix can be used (see, e.g., [13]). These methods have in common that they do not consider the higher order moments. One exception is the recent paper by Boudt et al. [4] proposing to use information criteria based on nearest distance estimation between the factor-model implied higher order co-moments and their sample version. However, their method is not applicable to our setup in which the factor model exposures are estimated in a first step using regression.

To fill this gap, we introduce a selection method that determines the number of required factors for which a pre-specified set of co-moments of the unexplained return variation are jointly equal to zero. The null hypothesis reflects the key property of a valid factor model that, because of the independence of the idiosyncratic return variation, many of its co-moments are zero. In practice, the estimated higher order co-moments are never exactly zero because of sampling variability. One thus needs a test to conclude whether they are significantly different from zero and thus reject correct specification of the factor model. In this paper, we exploit this intuition and propose to select the factor model using Wald and Gumbel tests for the higher order co-moments of the return variation that cannot be explained by the factor exposure.

The proposed testing procedure requires to first estimate the factor model by ordinary least squares(OLS). The sample co-moments of the residuals are then used as an input for the test statistic. We prove the asymptotic distribution of the proposed Wald and Gumbel test statistics, and use Monte Carlo simulations to verify that the test also has good size and power in finite samples. Based on this analysis, we recommend to use the proposed Gumbel test for realistic setups with more than 15 assets. In an application to stock returns, we find that single factor market model is most often sufficient to describe the higher order co-moments of stock returns.

The rest of the article is organized as follows. Section 2 describes how to construct co-moments under the linear factor model and to evaluate the estimation impact of the factor model under misspecification. Section 3 introduces the test and derives the asymptotic distribution of the proposed test statistics. The finite sample size and power of the test are then evaluated through a Monte Carlo study in Section 4. The empirical application in Section 5 shows how the test can be used for factor model selection on stock returns. Finally, Section 6 concludes the paper by summarizing the main findings. The proofs are given in the Appendix.

2. Higher order co-moments and the factor model

In this section, we introduce the notation for the co-moments of interest in both the general case and the specific case of a factor model. We focus on an N-dimensional vector of asset returns, but the notation holds for any multivariate random vector.

2.1. Higher order co-moments

The parameters of interest are the co-moments of an N-dimensional vector of asset returns $\boldsymbol{R_t} = (R_{1t}, \dots, R_{Nt})'$, observed at a regular frequency ($t = 1, \dots, T$). We denote the first (uncentered) moment by μ_R. Then, the $N \times N$ covariance matrix Σ, $N \times N^2$ coskewness matrix $\boldsymbol{\Phi}$ and $N \times N^3$ cokurtosis matrix $\boldsymbol{\Psi}$ of the return vector $\boldsymbol{R_t}$ equal the following:

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathbb{E}[(\boldsymbol{R_t} - \mu_R)(\boldsymbol{R_t} - \mu_R)'],\\
\boldsymbol{\Phi} &= \mathbb{E}[(\boldsymbol{R_t} - \mu_R)(\boldsymbol{R_t} - \mu_R)' \otimes (\boldsymbol{R_t} - \mu_R)'],\\
\boldsymbol{\Psi} &= \mathbb{E}[(\boldsymbol{R_t} - \mu_R)(\boldsymbol{R_t} - \mu_R)' \otimes (\boldsymbol{R_t} - \mu_R)' \otimes (\boldsymbol{R_t} - \mu_R)'].
\end{aligned} \tag{1}
$$

Here, $\otimes$ denotes the Kronecker product. These definitions of higher order co-moments based on tensor approach are widespread (see, e.g., [5,14]), because they lend themselves to a computationally convenient expression of the moments of portfolio returns as explicit functions of the weight vector and the co-moment matrix. The N-dimensional portfolio weight vector is denoted by $\boldsymbol{w}$. The portfolio's second, third, and fourth moments are given by the following:

$$
\begin{aligned}
m_2(\boldsymbol{w}) &= \mathbb{E}[(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))'] = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w},\\
m_3(\boldsymbol{w}) &= \mathbb{E}[(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))' \otimes (\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))'] = \boldsymbol{w}'\boldsymbol{\Phi}(\boldsymbol{w} \otimes \boldsymbol{w}),\\
m_4(\boldsymbol{w}) &= \mathbb{E}[(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))(\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))' \otimes (\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))' \otimes (\boldsymbol{w}'(\boldsymbol{R_t} - \mu_R))']\\
&= \boldsymbol{w}'\boldsymbol{\Psi}(\boldsymbol{w} \otimes \boldsymbol{w} \otimes \boldsymbol{w}).
\end{aligned} \tag{2}
$$

2.2. Co-moments under the linear factor model

As recommended by Martellini and Ziemann [6] and Boudt et al. [7], we focus on the estimation of $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$ under the assumption that the asset returns are generated by a linear factor model with K factors $\boldsymbol{F}_t = (F_{1t}, \ldots, F_{Kt})'$, whereby the variation in the asset returns not explained by the factors, i.e., the error term, is assumed to be independent of each factor and across assets. A multi-factor linear model for N individual asset returns is given by the following:

$$R_{it} = a_i + \boldsymbol{\beta}_i'\boldsymbol{F}_t + \varepsilon_{it},$$
$$\varepsilon_{it} = \sigma_i \eta_{it}. \tag{3}$$

Here, ε_{it} is the idiosyncratic error term of asset i with standard deviation σ_i, η_{it} is the standardized idiosyncratic error term of asset i, and $\boldsymbol{\beta}_i$ is a vector of factor loadings associated with $\boldsymbol{F}_t$ (also called factor beta's or factor exposures). Then, the covariance matrix can be written as follows:

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{F}}\mathbf{B}' + \boldsymbol{\Delta}\boldsymbol{\Delta}. \tag{4}$$

Here, $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)'$ indicates the $N \times K$ matrix of regression parameters in Equation (3), $\boldsymbol{\Sigma}_F$ is the covariance matrix of the factors, and $\boldsymbol{\Delta}$ is the diagonal matrix with ith element equal to σ_i. Similar decompositions can be made for the coskewness matrix and cokurtosis matrix:

$$\boldsymbol{\Phi} = \mathbf{B}\boldsymbol{\Phi}_{\mathbf{F}}(\mathbf{B} \otimes \mathbf{B}') + \boldsymbol{\Omega},$$
$$\boldsymbol{\Psi} = \mathbf{B}\boldsymbol{\Psi}_{\mathbf{F}}(\mathbf{B}' \otimes \mathbf{B}' \otimes \mathbf{B}') + \boldsymbol{\Upsilon}. \tag{5}$$

Here, $\boldsymbol{\Phi}_F$ and $\boldsymbol{\Psi}_F$ are the coskewness and cokurtosis matrices of the factors, as defined in Equation (1), and we refer to Boudt et al. [7] for the exact expressions for the residual matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Upsilon}$ containing mostly zeros.

3. Testing sparsity of higher order co-moments

An important property of the factor model is that, if there are no important variables omitted in the model, then the error terms should be independent across each other. In contrast, if there is model misspecification (due to e.g., the omission of relevant factors), this typically leads to some form of dependence in the error terms. We will exploit this feature to construct a diagnostic test for appropriate model specification.

The model diagnostic tests the independence of the error terms based on the estimated standardized residuals. In order to avoid nuisance parameters in the asymptotic distribution of the test statistic, we limit ourselves to test statistics for which the asymptotic distribution of the test statistic does not depend on the distribution of the first step regression estimates. As we show later, this is the case for the bi-product, tri-product, and quad-product of standardized error terms of different assets.

For the covariation between the unexplained return variation, the corresponding null hypotheses to be tested are as follows:

$$H1_0 : \mathbb{E}[\eta_{it}\eta_{jt}] = 0, \quad \forall\, i,j : i < j, \quad \forall\, t. \tag{6}$$

For the coskewness terms of the error, we should test the following:

$$H2_0 : \mathbb{E}[\eta_{it}\eta_{jt}\eta_{kt}] = 0, \quad \forall\, i,j,k : i < j < k, \quad \forall\, t. \tag{7}$$

Finally, for the cokurtosis terms of the error, we should test the following:

$$H3_0 : \mathbb{E}[\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}] = 0, \quad \forall\, i,j,k,l : i < j < k < l, \quad \forall\, t. \tag{8}$$

3.1. Distribution of the bi-product, tri-product, and quad-product of the idiosyncratic shocks of different assets under correct specification

First, $\boldsymbol{\eta}_{[2]t}$ denotes the $P_2 \times 1$ bi-product vector containing $\eta_{it} \cdot \eta_{jt}$, with $i < j$ and $P_2 = N(N-1)/2$, i.e.,

$$\boldsymbol{\eta}_{[2]t} = (\eta_{1t}\eta_{2t}, \dots, \eta_{1t}\eta_{Nt}, \eta_{2t}\eta_{3t}, \dots, \eta_{N-1t}\eta_{Nt})'.$$

Second, $\boldsymbol{\eta}_{[3]t}$ denotes the $P_3 \times 1$ tri-product vector containing $\eta_{it} \cdot \eta_{jt} \cdot \eta_{kt}$, with $i < j < k$ and $P_3 = N(N-1)(N-2)/6$, i.e.,

$$\boldsymbol{\eta}_{[3]t} = (\eta_{1t}\eta_{2t}\eta_{3t}, \dots, \eta_{1t}\eta_{2t}\eta_{Nt}, \eta_{1t}\eta_{3t}\eta_{4t}, \dots, \eta_{N-2t}\eta_{N-1t}\eta_{Nt})'.$$

Finally, $\boldsymbol{\eta}_{[4]t}$ denotes the $P_4 \times 1$ quad-product vector containing $\eta_{it} \cdot \eta_{jt} \cdot \eta_{kt} \cdot \eta_{lt}$, with $i < j < k < l$ and $P_4 = N(N-1)(N-2)(N-3)/24$, i.e.,

$$\boldsymbol{\eta}_{[4]t} = (\eta_{1t}\eta_{2t}\eta_{3t}\eta_{4t}, \dots, \eta_{1t}\eta_{2t}\eta_{3t}\eta_{Nt}, \eta_{1t}\eta_{3t}\eta_{4t}\eta_{5t}, \dots, \eta_{N-3t}\eta_{N-2t}\eta_{N-1t}\eta_{Nt})'.$$

Note that, under the assumption that the returns are generated by the factor model in Equation (3), $\boldsymbol{\eta}_{[2]t}$, $\boldsymbol{\eta}_{[3]t}$, and $\boldsymbol{\eta}_{[4]t}$ all have zero mean and their covariance matrices are the identity matrix. By the central limit theorem, we have, if the returns are generated by the factor model in Equation (3), the following distributional results for the average bi-power, tri-power, and quad-power statistics of independent components:

$$\begin{aligned}
\boldsymbol{\eta}_{[2]} &= \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\eta}_{[2]t} \xrightarrow{d} N_{P_2}(\mathbf{0}; I_{P_2}/T), \\
\boldsymbol{\eta}_{[3]} &= \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\eta}_{[3]t} \xrightarrow{d} N_{P_3}(\mathbf{0}; I_{P_3}/T), \\
\boldsymbol{\eta}_{[4]} &= \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\eta}_{[4]t} \xrightarrow{d} N_{P_4}(\mathbf{0}; I_{P_4}/T).
\end{aligned} \tag{9}$$

3.2. Distribution of the estimated bi-product, tri-product, and quad-product of the idiosyncratic shocks of different assets under correct specification

To test the null hypothesis $H1_0$, $H2_0$, and $H3_0$ in (6)–(8), we need to aggregate across time and across asset combinations. We first stack all the bi-products, tri-products, and quad-products of the unexplained return variation into vectors. In practice, it is infeasible to compute $\boldsymbol{\eta}_{[2]}$, $\boldsymbol{\eta}_{[3]}$, and $\boldsymbol{\eta}_{[4]}$. Obtaining a feasible version requires to first estimate the factor model parameters. We estimate the parameters for each stock separately by OLS regression of the time-series regression model obtained by stacking observations for a given asset i:

$$\boldsymbol{R}_i = \mathbf{1}_T\alpha_i + \mathbf{F}'\boldsymbol{\beta}_{\boldsymbol{i}} + \boldsymbol{\varepsilon}_{\boldsymbol{i}}. \tag{10}$$

Here, $\boldsymbol{R}_{\boldsymbol{i}} = [R_{i1}, R_{i2}, \ldots, R_{iT}]'$, $\mathbf{F} = [\boldsymbol{f}_{\mathbf{1}}, \boldsymbol{f}_{\mathbf{2}}, \ldots, \boldsymbol{f}_{\boldsymbol{K}}]'$ is the $(K \times T)$ matrix of the observed values of the K factors and $\boldsymbol{\varepsilon}_{\boldsymbol{i}} = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}]'$.

When Equation (10) is estimated using the correct set of factors, then asymptotically, the residuals will not display any higher order dependence. In contrast, if relevant factors are omitted, this may induce higher order dependence in the residual series. We next exploit this result to propose our diagnostic tool for correct factor model specification based on testing the sparsity of the higher order moments of the residuals.

The corresponding estimated standardized residuals are denoted through the following system:

$$\hat{\boldsymbol{\eta}}_{\boldsymbol{t}} = \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{r}_{\boldsymbol{t}} - \hat{\boldsymbol{a}} - \hat{\mathbf{B}}\boldsymbol{F}_t). \tag{11}$$

Here, $\hat{\boldsymbol{a}}$ and $\hat{\mathbf{B}}$ are the OLS estimates of the intercept and factor exposure parameters, and $\hat{\boldsymbol{\Delta}}$ contains the estimated standard deviation of the regression errors.

We then construct the estimated bi-power, tri-power and quad-power statistics:

$$\hat{\boldsymbol{\eta}}_{[2]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}, \quad \hat{\boldsymbol{\eta}}_{[3]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}}, \quad \hat{\boldsymbol{\eta}}_{[4]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[4]\boldsymbol{t}}. \tag{12}$$

Here, $\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}$ is the $P_2 \times 1$ bi-product vector containing $\hat{\eta}_{it} \cdot \hat{\eta}_{jt}$ with $i < j$; $\hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}}$ is the $P_3 \times 1$ tri-product vector containing $\hat{\eta}_{it} \cdot \hat{\eta}_{jt} \cdot \hat{\eta}_{kt}$ with $i < j < k$ and $\hat{\boldsymbol{\eta}}_{[4]\boldsymbol{t}}$ is the $P_4 \times 1$ quad-product vector containing $\hat{\eta}_{it} \cdot \hat{\eta}_{jt} \cdot \hat{\eta}_{kt} \cdot \hat{\eta}_{lt}$ with $i < j < k < l$. If we combine the covariance and coskewness terms, we then construct the estimated bi-power plus tri-power statistics:

$$\hat{\boldsymbol{\eta}}_{[5]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[5]\boldsymbol{t}}. \tag{13}$$

Here, $\hat{\boldsymbol{\eta}}_{[5]\boldsymbol{t}} = (\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}})'$. If we combine all higher order co-moment terms, we then obtain the joint estimated bi-power, tri-power and quad-power as follows:

$$\hat{\boldsymbol{\eta}}_{[6]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[6]\boldsymbol{t}}. \tag{14}$$

Here, $\hat{\boldsymbol{\eta}}_{[6]\boldsymbol{t}} = (\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[4]\boldsymbol{t}})'$.

We will turn the observed vectors $\hat{\boldsymbol{\eta}}_{[2]}, \hat{\boldsymbol{\eta}}_{[3]}, \hat{\boldsymbol{\eta}}_{[4]}, \hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ into scalar-valued Wald and Gumbel test statistics. Before showing the explicit forms of the test statistics, we first derive conditions under which $\hat{\boldsymbol{\eta}}_{[2]}, \hat{\boldsymbol{\eta}}_{[3]}, \hat{\boldsymbol{\eta}}_{[4]}, \hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ converge in distribution to $\boldsymbol{\eta}_{[2]}, \boldsymbol{\eta}_{[3]}, \boldsymbol{\eta}_{[4]}, \boldsymbol{\eta}_{[5]}$, and $\boldsymbol{\eta}_{[6]}$, and are thus asymptotically multivariate normally distribution around $\mathbf{0}$ and with the identity matrix as the covariance matrix.

Let $\boldsymbol{b} = (\boldsymbol{a}, \mathrm{vec}(\mathbf{B}), \sigma_1, \ldots, \sigma_N)'$ be the parameter vector, and $\hat{\boldsymbol{b}}$ the corresponding estimate. When $\hat{\boldsymbol{\eta}}_{[2]}, \hat{\boldsymbol{\eta}}_{[3]}, \hat{\boldsymbol{\eta}}_{[4]}, \hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ are evaluated at $\boldsymbol{b}$, and $\hat{\boldsymbol{\eta}}_{[2]}, \hat{\boldsymbol{\eta}}_{[3]}, \hat{\boldsymbol{\eta}}_{[4]}, \hat{\boldsymbol{\eta}}_{[5]}$, and

$\hat{\boldsymbol{\eta}}_{[6]}$ of course coincide with $\boldsymbol{\eta}_{[2]}$, $\boldsymbol{\eta}_{[3]}$, $\boldsymbol{\eta}_{[4]}$, $\boldsymbol{\eta}_{[5]}$, and $\boldsymbol{\eta}_{[6]}$. Since these are U-statistics, we can follow Randles [15] in proving that, the asymptotic distribution of the test statistics $\hat{\boldsymbol{\eta}}_{[2]}$, $\hat{\boldsymbol{\eta}}_{[3]}$, $\hat{\boldsymbol{\eta}}_{[4]}$, $\hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ are independent of the estimation uncertainty in the factor exposures, provided the following first order approximation holds.

Assumption 3.1: The statistic $\hat{\boldsymbol{\eta}}_{[j]}, j = 2, 3, 4, 5, 6$ can be approximated by the following:

$$\sqrt{T}\hat{\boldsymbol{\eta}}_{[j]} = \sqrt{T}\boldsymbol{\eta}_{[j]} + \boldsymbol{\mu}_{[j]}\sqrt{T}(\hat{\boldsymbol{b}} - \boldsymbol{b}) + o_p(\mathbf{1}), \tag{15}$$

Here, $\boldsymbol{\eta}_{[j]}$ is differentiable in $\boldsymbol{b}$, $\hat{\boldsymbol{b}}$ is a consistent estimator of $\boldsymbol{b}$, and $\boldsymbol{\mu}_{[j]} = \lim_{T\to\infty} E[\partial\boldsymbol{\eta}_{[j]}/\partial\boldsymbol{b}]$.

Note that, under Assumption 3.1, the asymptotic distribution of $\sqrt{T}\hat{\boldsymbol{\eta}}_{[j]}$ and $\sqrt{T}\boldsymbol{\eta}_{[j]}$ coincide when $\boldsymbol{\mu}_{[j]} = 0$. In Appendix, we prove that this is the case, since the limiting value of the mean of the gradient of $\boldsymbol{\eta}_{[2]}$, $\boldsymbol{\eta}_{[3]}$, $\boldsymbol{\eta}_{[4]}$, $\boldsymbol{\eta}_{[5]}$, and $\boldsymbol{\eta}_{[6]}$ with respect to all the the parameters in the factor model (3), including the intercept $\boldsymbol{a}$, the factor exposure, and the variance of the specific factors, is zero:

$$\begin{aligned} \lim_{T\to\infty} \mathbb{E}\left[\frac{\partial\boldsymbol{\eta}_{[2]}}{\partial\boldsymbol{b}}\right] &= \lim_{T\to\infty} \mathbb{E}\left[\frac{\partial\boldsymbol{\eta}_{[3]}}{\partial\boldsymbol{b}}\right] = \lim_{T\to\infty} \mathbb{E}\left[\frac{\partial\boldsymbol{\eta}_{[4]}}{\partial\boldsymbol{b}}\right] = \lim_{T\to\infty} \mathbb{E}\left[\frac{\partial\boldsymbol{\eta}_{[5]}}{\partial\boldsymbol{b}}\right] \\ &= \lim_{T\to\infty} \mathbb{E}\left[\frac{\partial\boldsymbol{\eta}_{[6]}}{\partial\boldsymbol{b}}\right] = \mathbf{0}. \end{aligned} \tag{16}$$

The proposition below then follows directly.

Proposition 3.1: *When returns are generated according to the factor model in (3) and Assumption 3.1 holds, then $\sqrt{T}\hat{\boldsymbol{\eta}}_{[2]}, \sqrt{T}\hat{\boldsymbol{\eta}}_{[3]}, \sqrt{T}\hat{\boldsymbol{\eta}}_{[4]}, \sqrt{T}\hat{\boldsymbol{\eta}}_{[5]}$ and $\sqrt{T}\hat{\boldsymbol{\eta}}_{[6]}$ defined in (13) and (14) are asymptotically multivariate normal distributed with zero mean and the identity matrix as covariance matrix.*

3.3. Wald and Gumbel tests for detecting misspecification of the factor model

The above results can be used to test the sparsity assumptions in (6)–(8) in various ways. We next present two types of tests, namely a Wald test and a Gumbel test. We leave it for further research to consider other types of test statistics exploiting the null distribution presented in Proposition 3.1.

The Wald test evaluates whether the mean value of the covariance elements of $\boldsymbol{\eta}_{[2]}$, the coskewness elements of $\boldsymbol{\eta}_{[3]}$, and the cokurtosis elements of $\boldsymbol{\eta}_{[4]}$ are zero based on their estimated values $\hat{\boldsymbol{\eta}}_{[2]}$, $\hat{\boldsymbol{\eta}}_{[3]}$, and $\hat{\boldsymbol{\eta}}_{[4]}$. It follows from Proposition 3.1 that, under the null hypothesis of correct specification, the Wald test statistic obtained as a scaled sum of the squared elements in $\hat{\boldsymbol{\eta}}_{[2]}$, $\hat{\boldsymbol{\eta}}_{[3]}$, $\hat{\boldsymbol{\eta}}_{[4]}$, $\hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ asymptotically converge to a chi-square distribution with P_2, P_3, P_4, $P_2 + P_3$, and $P_2 + P_3 + P_4$ degrees of freedom, respectively,

$$W_{[2]} = T\hat{\boldsymbol{\eta}}'_{[2]}\hat{\boldsymbol{\eta}}_{[2]} \xrightarrow{d} \chi^2_{P_2},$$

$$W_{[3]} = T\hat{\boldsymbol{\eta}}'_{[3]}\hat{\boldsymbol{\eta}}_{[3]} \xrightarrow{d} \chi^2_{P_3},$$

$$W_{[4]} = T\hat{\boldsymbol{\eta}}'_{[4]}\hat{\boldsymbol{\eta}}_{[4]} \xrightarrow{d} \chi^2_{P_4},$$
$$W_{[5]} = T\hat{\boldsymbol{\eta}}'_{[5]}\hat{\boldsymbol{\eta}}_{[5]} \xrightarrow{d} \chi^2_{P_2+P_3},$$
$$W_{[6]} = T\hat{\boldsymbol{\eta}}'_{[6]}\hat{\boldsymbol{\eta}}_{[6]} \xrightarrow{d} \chi^2_{P_2+P_3+P_4}, \tag{17}$$

for $T \to \infty$. These test statistics, and their associated p-values, are simple and fast to compute using the co-moment functions of the R package PerformanceAnalytics (see, e.g., [16,17]).

The proposed Gumbel test uses a maximum statistic and a Gaussian score transformation of the residual variation in $\hat{\boldsymbol{\eta}}_t$ in (11). Specifically, $\hat{\boldsymbol{k}}_t$ denotes the vector of ranks of the corresponding elements in $\hat{\boldsymbol{\eta}}_t$. The corresponding Gaussian (or normal) scores are obtained by plugging these ranks in the quantile function Φ^{-1} of the standard normal distribution:

$$\hat{\boldsymbol{\zeta}}_t = \Phi^{-1}\left(\frac{\hat{\boldsymbol{k}}_t}{T+1}\right) \tag{18}$$

Note that this transformation yields a series that has no univariate outliers, but still preserves the dependence characteristics of the original series. Such a transformation of data to the Gaussian scores, also called the Van Der Waerden scores or the normal scores, has been used before to obtain robust correlation estimators (see e.g., [18–20]). The robustness follows from the use of ranks. As before, we can then compute the corresponding estimated bi-power, tri-power, and quad-power statistics:

$$\hat{\boldsymbol{\zeta}}_{[2]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\zeta}}_{[2]\boldsymbol{t}}, \quad \hat{\boldsymbol{\zeta}}_{[3]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\zeta}}_{[3]\boldsymbol{t}}, \quad \hat{\boldsymbol{\zeta}}_{[4]} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\zeta}}_{[4]\boldsymbol{t}}. \tag{19}$$

Here, $\hat{\boldsymbol{\zeta}}_{[2]}$ is the $P_2 \times 1$ bi-product vector containing $\hat{\zeta}_{it} \cdot \hat{\zeta}_{jt}$ with $i < j$; $\hat{\boldsymbol{\zeta}}_{[3]\boldsymbol{t}}$ is the $P_3 \times 1$ tri-product vector containing $\hat{\zeta}_{it} \cdot \hat{\zeta}_{jt} \cdot \hat{\zeta}_{kt}$ with $i < j < k$; and $\hat{\boldsymbol{\eta}}_{[4]\boldsymbol{t}}$ is the $P_4 \times 1$ quad-product vector containing $\hat{\zeta}_{it} \cdot \hat{\zeta}_{jt} \cdot \hat{\zeta}_{kt} \cdot \hat{\zeta}_{lt}$ with $i < j < k < l$.

The Gumbel test, recommended for medium to high dimensions, uses the property of the Gumbel distribution as the extreme value distribution for the maximum of a set of absolute values of normally distributed random variables. The corresponding test statistics are as follows:

$$G_{[2]} = \frac{\max_i |\hat{\boldsymbol{\zeta}}_{[2]\boldsymbol{i}}| - C_{P_2}}{S_{P_2}} \overset{a}{\sim} Gumbel,$$
$$G_{[3]} = \frac{\max_i |\hat{\boldsymbol{\zeta}}_{[3]\boldsymbol{i}}| - C_{P_3}}{S_{P_3}} \overset{a}{\sim} Gumbel,$$
$$G_{[4]} = \frac{\max_i |\hat{\boldsymbol{\zeta}}_{[4]\boldsymbol{i}}| - C_{P_4}}{S_{P_4}} \overset{a}{\sim} Gumbel,$$
$$G_{[5]} = \frac{\max_i |\hat{\boldsymbol{\zeta}}_{[5]\boldsymbol{i}}| - C_{P_2+P_3}}{S_{P_2+P_3}} \overset{a}{\sim} Gumbel,$$

$$G_{[6]} = \frac{\max_i |\hat{\boldsymbol{\xi}}_{[6]i}| - C_{P_2+P_3+P_4}}{S_{P_2+P_3+P_4}} \overset{a}{\sim} Gumbel. \tag{20}$$

Here, $C_n = (2\log n)^{0.5} - (\log(\pi) + \log(\log n))/2(2\log n)^{0.5}$ and $S_n = 1/(2\log n)^{0.5}$, n being the total number of elements over which the maximum is taken. The corresponding critical value for a probability α of type I error is given by $-\log(-\log(1-\alpha))$.

3.4. A general to specific approach to choosing the number of factors

In many cases, a large number of potentially relevant factors exist. We recommend to first estimate the higher order co-moments with all factors and to then reduce sequentially the set of factors by omitting those for which the Wald or Gumbel test do not reject the null hypothesis that the higher order co-moments of the idiosyncratic terms are zero. Such a general to specific model selection approach is widespread in regression analysis. In the simulation study and empirical application, we show how this approach can also be used for selecting the number of factors when the objective is to estimate the higher order co-moments. We illustrate this idea next in the simulation study and empirical application.

4. Monte Carlo Study

In this section, we use a simulation study to evaluate the finite sample properties of the test statistics. We consider a setup with at most five factors $\boldsymbol{F_i}$, where the models considered differ in terms of factors included and where, if factor $\boldsymbol{F_j}$ is included, all factors $\boldsymbol{F_i}$ with $i < j$ are also included. We use K_0 to denote the true number of factors and K_a to denote the assumed number of factors. For $K_0 = K_a$, the model is thus correctly specified, and we evaluate the size of the proposed sparsity tests. For $K_a < K_0$, we evaluate the power of the test. We set the dimension N either to 5, 15, or 30, and let T be such that T/N is either 10, 50, 250, or 1000. The simulation model assumes the innovations to be i.i.d. skewed Student t distributed.

The factor model is calibrated on daily returns of 30 large capitalization US stocks from January 2001 to December 2016, using the five factors proposed by Fama and French [10], namely market, size, value, profitability, and investment.[1] Stock returns are computed on adjusted close prices retrieved from Yahoo!Finance, while the factor data are downloaded from Kenneth French's website. Summary statistics for the daily returns of these 30 assets and 5 factors are reported in Table 1. They indicate that both stock returns and factors have a non-normal distribution and are thus a relevant setup to analyze the properties of the proposed Wald and Gumbel test.

Below, Section 4.1 describes in more detail the procedure to simulate the data. Subsequently, Section 4.2 reports the simulation results.

4.1. Simulation setup

For studying the size of the test, we carry out the following steps:

Step 1. Assume that K_0 is the true number of factors in the factor model generating the stock returns. Draw with replacement a random sample of $\boldsymbol{F_t^{K_0}} = (F_{1t}, \ldots, F_{K_0 t})'$.

Table 1. Summary statistics for the daily returns of 30 assets, together with the sample means and sample covariance matrix of $\boldsymbol{f}_t$ from 1/1/2000 to 12/31/2016.

Ticker	Min	Mean	Max	Std	Skew	Kur	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$	$\hat{\xi}$
AAPL	−19.747	0.117	13.019	2.390	−0.148	8.360	0.120	0.025	3.474	1.022
AIG	−93.626	−0.085	50.682	4.101	−3.241	107.495	−0.001	0.025	3.000	0.995
AMZN	−28.457	0.096	29.618	3.179	0.603	17.515	0.074	0.031	3.000	1.020
BAC	−34.206	−0.001	30.210	3.054	−0.359	29.740	−0.009	0.024	3.000	0.976
BIIB	−55.432	0.037	18.454	2.735	−2.772	57.504	0.043	0.023	9.989	1.015
B	−31.687	−0.005	22.159	2.422	−0.217	22.531	−0.001	0.022	3.000	0.986
C	−49.470	−0.053	45.632	3.268	−0.532	42.534	−0.001	0.025	3.000	0.989
CELG	−17.722	0.083	18.987	2.742	−0.092	9.410	0.057	0.029	3.144	0.988
CL	−11.910	0.018	9.986	1.245	−0.174	11.173	0.001	0.011	9.614	0.950
COF	−50.688	0.007	23.452	3.125	−1.383	31.007	0.001	0.028	3.000	1.001
CSCO	−17.686	−0.006	21.824	2.405	0.109	13.204	−0.001	0.024	3.000	0.965
DUK	−16.140	0.001	14.978	1.574	−0.277	15.67	−0.001	0.016	3.000	0.934
GILD	−16.038	0.082	12.949	2.399	−0.222	8.639	0.089	0.026	3.137	1.009
GS	−21.022	0.020	23.482	2.311	0.308	17.028	0.001	0.024	3.000	0.978
INTC	−20.479	0.005	18.335	2.243	−0.235	10.708	0.001	0.024	3.081	0.978
JPM	−23.228	0.016	22.392	2.542	0.269	17.308	0.001	0.023	3.000	0.985
MET	−31.156	0.011	24.586	2.692	−0.389	25.305	0.107	0.024	3.000	1.018
MO	−25.244	0.040	15.165	1.490	−1.728	34.919	0.034	0.015	3.104	0.960
MRK	−31.171	−0.012	12.251	1.740	−1.747	35.994	−0.001	0.015	9.382	0.921
MS	−29.966	−0.016	62.585	3.321	1.390	52.638	−0.039	0.025	9.397	0.925
ORCL	−23.639	0.007	19.332	2.351	−0.173	11.941	0.000	0.024	3.000	0.974
OXY	−20.448	0.045	16.643	2.097	−0.271	12.609	0.000	0.019	9.327	0.913
PCLN	−50.750	0.130	30.657	3.798	0.223	21.216	0.115	0.033	3.000	1.038
PEP	−12.705	0.019	13.862	1.183	−0.010	17.339	0.000	0.011	9.535	0.940
QCOM	−16.547	0.011	17.117	2.450	0.059	9.392	0.000	0.022	9.996	1.005
SPG	−22.314	−0.051	22.595	2.243	0.223	21.216	0.001	0.020	3.000	0.914
TXN	−20.119	0.011	13.859	2.356	−0.085	8.056	0.000	0.026	3.000	0.975
USB	−20.047	0.020	20.572	2.162	−0.192	19.816	0.000	0.019	3.000	0.975
WFC	−27.210	0.017	28.341	2.475	0.925	31.329	0.000	0.020	3.000	0.995
XOM	−15.027	0.018	15.863	1.536	−0.016	14.409	0.000	0.015	3.615	0.951
Factors	Min	Mean	Max	Std	Skew	Kur	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$	$\hat{\xi}$
MKT	−8.950	0.026	11.350	1.236	−0.057	10.929	0.000	0.011	9.240	0.902
SMB	−3.410	0.018	4.480	0.575	0.014	6.197	0.017	0.006	9.873	0.974
HML	−4.220	0.013	4.830	0.638	0.427	13.213	0.000	0.005	9.980	1.025
RMW	−2.680	0.018	2.990	0.455	−0.000	6.295	0.013	0.004	9.979	1.027
CMA	−5.930	0.011	2.320	0.380	−1.075	22.980	0.000	0.004	9.967	1.043
								Cov_F		
						1.528	0.108	0.149	−0.244	−0.095
						0.108	0.331	0.038	−0.066	0.019
						0.149	0.038	0.407	−0.042	0.082
						−0.245	−0.066	0.042	0.207	0.021
						−0.095	0.019	0.082	0.021	0.145

Notes: Asset and factor returns are expressed in percentage points. $\hat{\mu}$ as the mode, models the location, $\hat{\sigma} > 0$ is the dispersion parameter, $\hat{\nu} > 0$ models the tail thickness, and $\hat{\xi} > 0$ models the skewness.

Step 2. Then, generate the error terms from the standardized skewed Student t distribution, as proposed by Fernández and Steel [21] and Lambert and Laurent [22], as fitted on the stock returns, and scale them using the standard deviations as obtained on the calibration sample.

Step 3. Then, from model (3) and the obtained calibration for the intercept and loading matrix, we get a random sample of $\boldsymbol{R_t}$ with size T and assets N for the assumed factor number K_0.

Step 4. Perform the regression of model (3) using the generated return sample $\boldsymbol{R_t}$ and factor sample $\boldsymbol{F_t^{K_0}}$ under the true factor number K_0 to get the estimation of the exposure coefficients and the standardized residual vector $\hat{\eta}_t$. Compute the test statistics and their p values in (17) and (20).

For studying the power of the test, we change step 4 as follows.

Step 4'. Perform the regression of model (3) using the generated return sample $\boldsymbol{R}_t$ and factor sample $\boldsymbol{F}_t^{K_a}$ under the false factor number K_a from $\boldsymbol{F}_t^{K_0}(K_a < K_0)$ to get the estimation of the exposure coefficients and the standardized residual vector $\hat{\eta}_t$. Compute the test statistics and their p values in (17) and (20).

Finally, we also evaluate the accuracy of the factor selection approach. We then test sequentially $K_a = 5$, $K_a = 3$ and $K_a = 1$ and stop the iterative testing when the selected number of factors is no longer consistent with the null hypothesis. The number of factors selected equals the lowest number of factors for which the null hypothesis is not rejected.

Table 2. Size of Wald and Gumbel tests when testing at a 10% significance level (replications $M = 1000$).

			$H1_0 : E[\eta_i\eta_j] = 0$, $i<j$		$H2_0 : E[\eta_i\eta_j\eta_k] = 0$, $i<j<k$		$H3_0 : E[\eta_i\eta_j\eta_k\eta_l] = 0$, $i<j<k<l$		$H1_0 + H2_0$		$H1_0 + H2_0 + H3_0$	
N	K_0	T/N	Wald	Gumbel	Wald	Gumbel	Wald	Gumbel	Wald	Gumbel	Wald	Gumbel
5	1	10	0.114	0.014	0.134	0.007	0.107	0.009	0.132	0.013	0.146	0.008
		50	0.110	0.045	0.112	0.035	0.120	0.024	0.122	0.041	0.148	0.042
		250	0.118	0.062	0.111	0.048	0.106	0.036	0.117	0.054	0.117	0.061
		1000	0.114	0.061	0.124	0.059	0.116	0.057	0.107	0.069	0.113	0.072
	3	10	0.130	0.015	0.130	0.011	0.118	0.005	0.157	0.011	0.163	0.009
		50	0.110	0.045	0.113	0.025	0.118	0.023	0.130	0.037	0.151	0.042
		250	0.113	0.060	0.111	0.047	0.109	0.043	0.118	0.057	0.120	0.062
		1000	0.116	0.057	0.119	0.060	0.114	0.057	0.106	0.071	0.115	0.071
	5	10	0.160	0.022	0.138	0.008	0.140	0.008	0.166	0.013	0.179	0.010
		50	0.114	0.050	0.116	0.028	0.122	0.023	0.136	0.040	0.156	0.042
		250	0.117	0.062	0.113	0.047	0.106	0.040	0.118	0.056	0.120	0.065
		1000	0.117	0.059	0.118	0.060	0.112	0.054	0.109	0.066	0.114	0.072
15	1	10	0.110	0.029	0.238	0.027	0.295	0.058	0.234	0.029	0.305	0.053
		50	0.098	0.056	0.198	0.072	0.271	0.080	0.203	0.072	0.279	0.072
		250	0.103	0.082	0.148	0.093	0.240	0.109	0.155	0.091	0.250	0.105
		1000	0.110	0.082	0.132	0.101	0.199	0.084	0.125	0.098	0.200	0.090
	3	10	0.126	0.027	0.238	0.029	0.302	0.065	0.250	0.027	0.310	0.062
		50	0.101	0.059	0.201	0.074	0.275	0.082	0.210	0.071	0.281	0.081
		250	0.102	0.081	0.144	0.088	0.243	0.115	0.145	0.093	0.249	0.100
		1000	0.109	0.084	0.134	0.100	0.194	0.082	0.130	0.096	0.197	0.090
	5	10	0.156	0.035	0.239	0.021	0.301	0.070	0.254	0.023	0.312	0.066
		50	0.102	0.060	0.200	0.073	0.273	0.077	0.207	0.073	0.286	0.080
		250	0.104	0.082	0.143	0.093	0.239	0.114	0.149	0.094	0.250	0.104
		1000	0.107	0.082	0.133	0.102	0.196	0.082	0.129	0.097	0.200	0.089
30	1	10	0.125	0.045	0.292	0.042	0.384	0.147	0.298	0.039	0.387	0.135
		50	0.118	0.088	0.218	0.066	0.334	0.113	0.214	0.062	0.331	0.109
		250	0.118	0.089	0.183	0.088	0.320	0.084	0.177	0.088	0.317	0.078
		1000	0.104	0.086	0.130	0.081	0.298	0.098	0.136	0.081	0.299	0.099
	3	10	0.148	0.053	0.285	0.033	0.391	0.126	0.294	0.035	0.390	0.121
		50	0.119	0.084	0.216	0.065	0.331	0.107	0.213	0.062	0.327	0.102
		250	0.118	0.092	0.186	0.089	0.311	0.079	0.179	0.089	0.317	0.074
		1000	0.101	0.086	0.130	0.079	0.306	0.094	0.133	0.080	0.304	0.099
	5	10	0.167	0.059	0.293	0.043	0.388	0.142	0.302	0.046	0.392	0.127
		50	0.122	0.085	0.210	0.071	0.330	0.105	0.212	0.068	0.327	0.106
		250	0.115	0.090	0.175	0.083	0.318	0.082	0.181	0.088	0.320	0.080
		1000	0.101	0.088	0.133	0.081	0.301	0.095	0.136	0.081	0.303	0.098

4.2. Results

Table 2 reports the results for the Monte Carlo evaluation of the size of the Wald and Gumbel tests when the significance level is equal to 10%. The main result is that, for $N = 5$, the Wald test has an empirical size that is close to the nominal level, while for $N = 15$ and $N = 30$, the Wald test has a serious size distortion for the null hypothesis involving cokurtosis elements. This is due to the large number of cokurtosis elements to be estimated. Since the Gumbel test is an extremum test statistic, it is expected that its size is incorrect for $N = 5$. However, for $N = 15$ and $N = 30$, its size is always close to the nominal size when $T/N \geq 250$. The results are similar for the different values of K_0 considered. Based on this size evaluation, we thus recommend to use the Gumbel test when N is large enough.

Table 3 reports the power of the Wald and Gumbel tests when testing using a significance level of 10%. We focus our discussion on the Gumbel test statistic, as it has satisfactory size properties for $N = 15$ and $N = 30$. We find that the Gumbel test lacks power

Table 3. Power of Wald and Gumbel tests when testing at a 10% significance level (replications $M = 1000$).

				$H1_0 : E[\eta_i\eta_j] = 0$, $i < j$		$H2_0 : E[\eta_i\eta_j\eta_k] = 0$, $i < j < k$		$H3_0 : E[\eta_i\eta_j\eta_k\eta_l] = 0$, $i < j < k < l$		$H1_0 + H2_0$		$H1_0 + H2_0 + H3_0$	
N	K_0	K_a	T/N	*Wald*	*Gumbel*	*Wald*	*Gumbel*	*Wald*	*Gumbel*	*Wald*	*Gumbel*	*Wald*	*Gumbel*
5	3	1	10	0.949	0.755	0.244	0.013	0.351	0.076	0.875	0.662	0.834	0.634
			50	1.000	1.000	0.304	0.073	0.562	0.358	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.296	0.084	0.910	0.873	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.324	0.116	1.000	1.000	1.000	1.000	1.000	1.000
	5	1	10	0.887	0.575	0.247	0.011	0.380	0.100	0.801	0.476	0.767	0.450
			50	1.000	1.000	0.359	0.080	0.712	0.560	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.547	0.107	0.994	0.993	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.741	0.129	1.000	1.000	1.000	1.000	1.000	1.000
	5	3	10	0.573	0.208	0.171	0.009	0.200	0.025	0.466	0.143	0.415	0.125
			50	0.991	0.986	0.236	0.057	0.317	0.146	0.979	0.978	0.966	0.973
			250	1.000	1.000	0.336	0.079	0.642	0.514	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.551	0.113	0.978	0.978	1.000	1.000	1.000	1.000
15	3	1	10	1.000	1.000	0.672	0.048	0.999	1.000	1.000	1.000	1.000	1.000
			50	1.000	1.000	0.788	0.119	1.000	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.837	0.188	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.884	0.170	1.000	1.000	1.000	1.000	1.000	1.000
	5	1	10	1.000	1.000	0.618	0.046	0.952	0.996	1.000	1.000	0.987	1.000
			50	1.000	1.000	0.750	0.134	1.000	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.955	0.204	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.991	0.298	1.000	1.000	1.000	1.000	1.000	1.000
	5	3	10	1.000	1.000	0.424	0.027	0.720	0.758	0.972	1.000	0.821	1.000
			50	1.000	1.000	0.536	0.113	0.990	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.801	0.148	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.974	0.325	1.000	1.000	1.000	1.000	1.000	1.000
30	3	1	10	1.000	1.000	0.898	0.088	1.000	1.000	1.000	1.000	1.000	1.000
			50	1.000	1.000	0.978	0.194	1.000	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.995	0.280	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	0.992	0.507	1.000	1.000	1.000	1.000	1.000	1.000
	5	1	10	1.000	1.000	0.890	0.087	1.000	1.000	1.000	1.000	1.000	1.000
			50	1.000	1.000	0.991	0.205	1.000	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	1.000	0.349	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	1.000	0.819	1.000	1.000	1.000	1.000	1.000	1.000
	5	3	10	1.000	1.000	0.730	0.082	0.998	1.000	1.000	1.000	0.999	1.000
			50	1.000	1.000	0.890	0.169	1.000	1.000	1.000	1.000	1.000	1.000
			250	1.000	1.000	0.983	0.302	1.000	1.000	1.000	1.000	1.000	1.000
			1000	1.000	1.000	1.000	0.724	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. Percentage of times that the Gumbel test selects the correct number of factors when return dimension is N, the number of observations equals T, and the true number of factors is given by K_0 (replications $M = 1000$).

N	K_0	T/N	$H1_0$	$H2_0$	$H3_0$	$H1_0 + H2_0$	$H1_0 + H2_0 + H3_0$
15	1	10	0.959	0.958	0.883	0.960	0.896
		50	0.935	0.909	0.891	0.910	0.898
		250	0.913	0.896	0.873	0.897	0.880
		1000	0.916	0.892	0.911	0.897	0.907
	3	10	0.962	0.042	0.905	0.966	0.909
		50	0.938	0.100	0.905	0.919	0.906
		250	0.917	0.156	0.878	0.902	0.887
		1000	0.916	0.134	0.914	0.899	0.908
	5	10	0.965	0.021	0.704	0.977	0.934
		50	0.940	0.087	0.923	0.927	0.920
		250	0.918	0.114	0.886	0.906	0.896
		1000	0.918	0.271	0.918	0.903	0.911
30	1	10	0.932	0.935	0.793	0.940	0.809
		50	0.906	0.916	0.860	0.922	0.869
		250	0.908	0.909	0.904	0.906	0.911
		1000	0.909	0.914	0.900	0.916	0.895
	3	10	0.936	0.080	0.829	0.950	0.840
		50	0.910	0.168	0.876	0.926	0.880
		250	0.908	0.247	0.910	0.909	0.914
		1000	0.911	0.463	0.903	0.918	0.896
	5	10	0.941	0.070	0.858	0.954	0.873
		50	0.915	0.144	0.895	0.932	0.894
		250	0.910	0.259	0.918	0.912	0.920
		1000	0.912	0.661	0.905	0.919	0.902

Notes: $H1_0 : E[\eta_{it}\eta_{jt}] = 0, i < j$. $H2_0 : E[\eta_{it}\eta_{jt}\eta_{kt}] = 0, i < j < k$. $H3_0 : E[\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}] = 0, i < j < k$.

Table 5. Percentage of selected number of factors in random samples of 15 and 30 S&P 100 stocks using the Gumbel test (replications $M = 1000$).

N	K	$H1_0$	$H2_0$	$H3_0$	$H1_0 + H2_0$	$H1_0 + H2_0 + H3_0$
15	1	0.802	0.776	0.774	0.774	0.769
	3	0.069	0.068	0.061	0.076	0.067
	5	0.051	0.072	0.069	0.068	0.076
	> 5	0.078	0.084	0.096	0.082	0.088
30	1	0.798	0.792	0.702	0.795	0.704
	3	0.060	0.070	0.097	0.067	0.099
	5	0.067	0.061	0.095	0.062	0.093
	> 5	0.075	0.077	0.106	0.076	0.104

Notes: $H1_0 : E[\eta_{it}\eta_{jt}] = 0, i < j$. $H2_0 : E[\eta_{it}\eta_{jt}\eta_{kt}] = 0, i < j < k$. $H3_0 : E[\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}] = 0, i < j < k$.

for the coskewness elements when T/N is small. Its power increases when T/N increases. The Gumbel test has perfect power in case it involves the covariance and/or cokurtosis elements.

Considering the Gumbel test has quite good size and power when T/N increases, Table 4 depicts the percentage of times the Gumbel test selects the correct number of factors when return dimension is N, the number of observations equals T, and the true number of factors is given by K_0. The simulation of the selection procedure is replicated 1000 times and the results are recorded.

From Table 4, we find that the Gumbel test tends to choose the correct number of factors most of the times, except for coskewness when T/N is small. The accuracy increases substantially when T/N becomes large. Thus, the Gumbel test is recommended in case $N = 15$ or larger and for large sample sizes.

5. Application

We apply the Gumbel test procedure to investigate whether the co-moments of stock returns are compatible with the single-factor model of Sharpe [23], the three-factor model of Fama and French [9], or the five-factor model of Fama and French [10]. Similar to the simulation study, we take the subset of $N = 15$ and $N = 30$ stocks, randomly chosen from the S&P 100 universe of stocks as of December 31, 2016. The sample period ranges from January 1, 2001 to December 31, 2016. We omit those with missing observations and end with a universe of 88 stocks from the S&P 100 database. A similar approach is adopted by Martellini and Ziemann [6]. We use daily returns and apply the general to specific factor selection procedure with a 10% significance level.

As an average over the randomly selected baskets of $N = 15$ and $N = 30$ stocks, Table 5 shows the percentage of selected number of factors using the proposed Gumbel test. We find that in most cases, a single-factor model is sufficient for explaining the co-moments.

6. Conclusion

In this paper, we propose a diagnostic tool to validate the use of factor models for estimating the higher order co-moments of asset returns. The factor model assumes the error terms to be independent, which implies sparsity of their co-moments. We propose a Wald test and a Gumbel test for evaluating this sparsity. We find that the Wald test is satisfactory in low dimensions, while the Gumbel test is recommended in case $N = 15$ or larger. In the empirical application to stock returns, we find that a single-factor model is most often detected as sufficient to estimate the co-moments.

We see several promising directions for further research. First, the proposed test statistic is sensitive to outliers in the data. An outlier-robust alternative is desirable, but it is challenging to construct, as it requires not only robustly estimating the factor model parameters but also defining a robust test statistic with a known distribution function under the null hypothesis. When the test statistic requires trimming, this is not straightforward. A second direction for further research is to allow for time-variation in the underlying factor model parameters. Finally, it would be important to evaluate in more detail the economic value of our test in a large scale empirical setting.

Note

1. The 30 tickers of the stocks are selected in terms of the largest total factor exposure from S&P 100 components. They are as follows: AAPL, AIG, AMZN, BAC, BIIB, BK, C, CELG, CL, COF, CSCO, DUK, GILD, GS, INTC, JPM, MET, MO, MRK, MS, ORCL, OXY, PCLN, PEP, QCOM, SPG, TXN, USB, WFC, and XOM.

Acknowledgments

We are grateful to two the anonymous referees and to the editor whose valuable suggestions and remarks allowed us to significantly improve the original paper. We also thank Dries Cornilly and Giang Nguyen for their constructive comments and assistance.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Wanbo Lu's research is sponsored by the National Science Foundation of China (71771187, 71101118), the Program for New Century Excellent Talents in University (NCET–13–0961), and the Fundamental Research Funds for the Central Universities (JBK170970) of China. Dong Yang's research is supported in part by China Scholarship Council, the Fundamental Research Funds for the Central Universities (JBK1807047, JBK1805004) of China, and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics.

References

[1] Boudt K, Peterson B, Croux C. Estimation and decomposition of downside risk for portfolios with non-normal returns. J Risk. 2008;11:79–103.
[2] Boudt K, Carl P, Peterson B. Asset allocation with conditional value-at-risk budgets. J Risk. 2013;15:39–68.
[3] Boudt K, Cornilly D, Verdonck T. A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. J Financ Econom, forthcoming.
[4] Boudt K, Cornilly D, Verdonck T. Nearest comoment estimation with unobserved factors. 2018. Available from SSRN: https://ssrn.com/abstract = 3087336.
[5] Jondeau E, Rockinger M. Optimal portfolio allocation under higher moments. Eur Financ Manag. 2006;12:29–55.
[6] Martellini L, Ziemann V. Improved estimates of higher-order comoments and implications for portfolio selection. Rev Financ Stud. 2010;23:1467–1502.
[7] Boudt K, Lu WB, Peeters B. Higher order comoments of multi factor models and asset allocation. Financ Res Lett. 2015;13:225–233.
[8] Brandt MW, Santalara P, Valkanov R. Parametric portfolio policies: exploiting characteristics in the cross-section of equity returns. Rev Financ Stud. 2009;22:3411–3447.
[9] Fama EF, French KR. Common risk factors in the returns on stocks and bonds. J Financ Econ. 1993;33:153–193.
[10] Fama EF, French KR. A five-factor asset pricing model. J Financ Econ. 2015;116:1–22.
[11] Cragg J, Donald S. Inferring the rank of a matrix. J Econom. 1997;76:223–250.
[12] Bai JS, Ng S. Determining the number of factors in approximate factor models. Econometrica. 2002;70:191–221.
[13] Ahn SC, Horenstein AR. Eigenvalue ratio test for the number of factors. Econometrica. 2013;81:1203–1227.
[14] Harvey CR, Liechty JC, Liechty MW, et al. Portfolio selection with higher moments. Quant Financ. 2010;10:469–485.
[15] Randles RH. On the asymptotic normality of statistics with estimated parameters. Ann Stat. 1982;10:462–474.
[16] Carl P, Peterson B, Performance analytics: econometric tools for performance and risk analysis. R package version 1.5.2.
[17] Cornilly D, Peterson B, Performance analytics estimation of higher order moments. Mimeo.
[18] Hájek J, Šidák Z, Sen PK. Theory of rank tests. San Diego (CA): Academic Press; 1967.
[19] Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. Commun Stat–Simul C. 1982;11:311–334.
[20] Boudt K, Cornelissen J, Croux C. The Gaussian rank correlation estimator: robustness properties. Stat Comput. 2012;22:471–483.
[21] Fernández C, Steel MFJ. On Bayesian modeling of fat tails and skewness. J Am Stat Assoc. 1998;93:359–371.

[22] Lambert P, Laurent S, Modelling financial time series using GARCH-type models and a skewed Student density. Stat Discussion; 2001; Belgium: Université Catholique de Louvain.
[23] Sharpe WF. A simplified model for portfolio analysis. Manag Sci. 1963;9:277–293.

Appendix. Proofs of the limiting value of the mean of the gradient for the test statistics

Here, we prove that the limiting value of the mean of the gradient of $\hat{\boldsymbol{\eta}}_{[2]}, \hat{\boldsymbol{\eta}}_{[3]}, \hat{\boldsymbol{\eta}}_{[4]}, \hat{\boldsymbol{\eta}}_{[5]}$, and $\hat{\boldsymbol{\eta}}_{[6]}$ with respect to all the parameters in the factor model (3), including the intercept $\boldsymbol{a}$, the factor exposure, and the variance of the specific factors, is zero:

$$\begin{aligned}\lim_{T\to\infty}\mathbb{E}\left[\frac{\partial \boldsymbol{\eta}_{[2]}}{\partial \boldsymbol{b}}\right] &= \lim_{T\to\infty}\mathbb{E}\left[\frac{\partial \boldsymbol{\eta}_{[3]}}{\partial \boldsymbol{b}}\right] = \lim_{T\to\infty}\mathbb{E}\left[\frac{\partial \boldsymbol{\eta}_{[4]}}{\partial \boldsymbol{b}}\right] \\ &= \lim_{T\to\infty}\mathbb{E}\left[\frac{\partial \boldsymbol{\eta}_{[5]}}{\partial \boldsymbol{b}}\right] = \lim_{T\to\infty}\mathbb{E}\left[\frac{\partial \boldsymbol{\eta}_{[6]}}{\partial \boldsymbol{b}}\right] = \mathbf{0}\end{aligned} \tag{A1}$$

For notational convenience and without loss of generality, we assume that the multi-factor models for asset returns have the general form $R_{it} = \alpha_i + \boldsymbol{b}_i'\boldsymbol{F}_t + \sigma_i\eta_{it}$.

For the covariance terms, taking the derivative of $\boldsymbol{\eta}_{[2]}$ with respect to α_i, $\boldsymbol{b}_i'$, and σ_i, we have the following:

$$\frac{\partial \boldsymbol{\eta}_{[2]}}{\partial \alpha_h} = \begin{cases} -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\eta_{jt}/\sigma_i, & h = i, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\eta_{it}/\sigma_j, & h = j, \\ 0, & h \neq i, j, \end{cases} \tag{A2}$$

$$\frac{\partial \boldsymbol{\eta}_{[2]}}{\partial \boldsymbol{b}_h'} = \begin{cases} -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{jt}/\sigma_i, & h = i, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{it}/\sigma_j, & h = j, \\ 0, & h \neq i, j, \end{cases} \tag{A3}$$

$$\frac{\partial \boldsymbol{\eta}_{[2]}}{\partial \sigma_h} = \begin{cases} -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\eta_{it}\eta_{jt}/\sigma_i, & h = i, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\eta_{it}\eta_{jt}/\sigma_j, & h = j, \\ 0, & h \neq i, j. \end{cases} \tag{A4}$$

Therefore, $\lim_{T\to\infty}\mathbb{E}[\partial\boldsymbol{\eta}_{[2]}/\partial\alpha_i] = 0$ and $\lim_{T\to\infty}\mathbb{E}[\partial\boldsymbol{\eta}_{[2]}/\partial\boldsymbol{b}_i'] = 0$ due to $\mathbb{E}[\eta_{it}] = 0, \mathbb{E}[\boldsymbol{F}_t\eta_{it}] = 0$ after taking iterative expectations and $\mathbb{E}[\eta_{it}\eta_{jt}] = 0$ for $\forall\, i \neq j$. Let $\boldsymbol{b} = (\boldsymbol{a}, \text{vec}(\mathbf{B}), (\sigma_1, \ldots, \sigma_N))'$, then $\lim_{T\to\infty}\mathbb{E}[\partial\boldsymbol{\eta}_{[2]}/\partial\boldsymbol{b}] = 0$.

For the coskewness terms, $\hat{\boldsymbol{\eta}}_{[3]} = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[3]t} = (1/T)\sum_{t=1}^{T}\hat{\eta}_{it}\hat{\eta}_{jt}\hat{\eta}_{kt}$ for $i<j<k$, then $\boldsymbol{\eta}_{[3]} = (1/T)\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}$. Thus, we have the following:

$$\frac{\partial \boldsymbol{\eta}_{[3]}}{\partial \alpha_h} = \begin{cases} -\frac{1}{T}\sum_{t=1}^{T}\eta_{jt}\eta_{kt}/\sigma_i, & h=i, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{kt}/\sigma_j, & h=j, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}/\sigma_k, & h=k, \\ 0, & h\neq i,j,k, \end{cases} \tag{A5}$$

$$\frac{\partial \boldsymbol{\eta}_{[3]}}{\partial \boldsymbol{b}'_h} = \begin{cases} -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{jt}\eta_{kt}/\sigma_i, & h=i, \\ -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{it}\eta_{kt}/\sigma_j, & h=j, \\ -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{it}\eta_{jt}/\sigma_j, & h=k, \\ 0, & h\neq i,j,k, \end{cases} \tag{A6}$$

$$\frac{\partial \boldsymbol{\eta}_{[3]}}{\partial \sigma_h} = \begin{cases} -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}/\sigma_i, & h=i, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}/\sigma_j, & h=j, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}/\sigma_k, & h=k, \\ 0, & h\neq i,j,k. \end{cases} \tag{A7}$$

Similarly, because $\mathbb{E}[\eta_{it}\eta_{jt}] = 0$, $\mathbb{E}[\boldsymbol{F}_t\eta_{it}\eta_{jt}] = 0$, and $\mathbb{E}[\eta_{it}\eta_{jt}\eta_{kt}] = 0$ for $\forall\, i\neq j\neq k$, we have $\lim_{T\to\infty}\mathbb{E}[\partial\boldsymbol{\eta}_{[3]}/\partial\boldsymbol{b}] = 0$.

For the cokurtosis terms, $\hat{\boldsymbol{\eta}}_{[4]} = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_{[4]t} = (1/T)\sum_{t=1}^{T}\hat{\eta}_{it}\hat{\eta}_{jt}\hat{\eta}_{kt}\hat{\eta}_{lt}$ for $i<j<k<l$, then $\boldsymbol{\eta}_{[4]} = (1/T)\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}$. Thus, we have the following:

$$\frac{\partial \boldsymbol{\eta}_{[4]}}{\partial \alpha_h} = \begin{cases} -\frac{1}{T}\sum_{t=1}^{T}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_i, & h=i, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{kt}\eta_{lt}/\sigma_j, & h=j, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{lt}/\sigma_k, & h=k, \\ -\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\eta_{jt}\eta_{kt}/\sigma_l, & h=l, \\ 0, & h\neq i,j,k,l, \end{cases} \tag{A8}$$

$$\frac{\partial \boldsymbol{\eta}_{[4]}}{\partial \boldsymbol{b}_h'} = \begin{cases} -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \boldsymbol{F_t}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_i, & h = i, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \boldsymbol{F_t}\eta_{it}\eta_{kt}\eta_{lt}/\sigma_j, & h = j, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \boldsymbol{F_t}\eta_{it}\eta_{jt}\eta_{lt}/\sigma_k, & h = k, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \boldsymbol{F_t}\eta_{it}\eta_{jt}\eta_{kt}/\sigma_l, & h = l, \\ 0, & h \neq i, j, k, \end{cases} \tag{A9}$$

$$\frac{\partial \boldsymbol{\eta}_{[4]}}{\partial \sigma_h} = \begin{cases} -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_i, & h = i, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_j, & h = j, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_k, & h = k, \\ -\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}/\sigma_l, & h = l, \\ 0, & h \neq i, j, k, l. \end{cases} \tag{A10}$$

Similarly, because $\mathbb{E}[\eta_{it}\eta_{jt}\eta_{kt}] = 0$, $\mathbb{E}[\boldsymbol{F_t}\eta_{it}\eta_{jt}\eta_{kt}] = 0$, and $\mathbb{E}[\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}] = 0$ for $\forall\, i \neq j \neq k \neq l$, we have $\lim_{T\to\infty} \mathbb{E}[\partial \boldsymbol{\eta}_{[4]}/\partial \boldsymbol{b}] = 0$.

Finally, if we combine these terms, $\hat{\boldsymbol{\eta}}_{[5]} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_{[5]\boldsymbol{t}}$ with $\hat{\boldsymbol{\eta}}_{[5]\boldsymbol{t}} = (\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}})'$ and $\hat{\boldsymbol{\eta}}_{[6]} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_{[6]\boldsymbol{t}}$ with $\hat{\boldsymbol{\eta}}_{[6]\boldsymbol{t}} = (\hat{\boldsymbol{\eta}}_{[2]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[3]\boldsymbol{t}}, \hat{\boldsymbol{\eta}}_{[4]\boldsymbol{t}})'$, we have $\lim_{T\to\infty} E[\partial \boldsymbol{\eta}_{[5]}/\partial \boldsymbol{b}] = \lim_{T\to\infty} \mathbb{E}[\partial \boldsymbol{\eta}_{[6]}/\partial \boldsymbol{b}] = 0$.